

README for Accessing Experimental Real-Time TRMM Multi-Satellite Precipitation Analysis (TMPA-RT) Data Sets

George J. Huffman

Laboratory for Atmospheres, NASA Goddard Space Flight Center
and Science Systems and Applications, Inc.

30 October 2007

News Archive

16 October 2007 Multiple system issues interfered with production, including the availability of trmmopen; the full RT data sets were posted by mid-afternoon.

11 July 2007 NESDIS inserted incorrect navigation was inserted into the processing stream for NOAA-16 during the period

1050/09 - 1219/10 July 2007

which caused all data to be displaced several thousand km down-track. Consequently, the NESDIS AMSU-B precipitation estimates for that period exhibit two errors. First, all estimates are mis-located. Second, the land/sea masking that the algorithm performs consequently worked on data from the wrong locations and passed through regions of extremely high artifacts in the high- and mid-latitude oceans, where icy/frozen surface was misinterpreted as oceanic precipitation. These errors appear in 3B40RT and 3B42RT for the time range

12Z/09 - 12Z/10 July 2007

in regions where the NOAA-16 estimates were not superseded by conical-scan microwave estimates.

6 April 2007 Mistaken network administration rules caused the near-total loss of input data from outside GSFC in the period late on 2 April 2007 to mid-afternoon on 4 April 2007. Offline, 3B41RT and 3B42RT have been recomputed for the affected period, but users should be aware that little microwave data are available in 3B42RT. The reprocessed files placed in the archive are:

3B41RT: 12Z 3 April 2007 to 14Z 4 April 2007

3B42RT: 12Z 3 April 2007 to 12Z 4 April 2007

At the same time, episodic networking failures at NCEP denied deliver of AMSU-B estimates for significant parts of 30 March to 5 April 2007, so the microwave content of 3B40RT and 3B42RT is reduced during that period. Finally, a power failure at NCEP interrupted AMSU-B and IR data delivery in the period 0330-0830Z on 6 April 2007. This failure again reduced the microwave content of 3B40RT and 3B42RT, reduced the IR content of 3B40RT at 03Z and 08Z and of 3B42RT at 03Z, and eliminated IR in the interval 04-07Z, so no 3B41RT and 3B42RT were produced for those times. Users can use the 06Z 3B40RT to fill in the missing 3B42RT (but note the different format).

6 April 2007 Mistakes in configuring network administration rules caused the near-total loss of input data from outside GSFC in the period late on Monday, 2 April to mid-afternoon on Wednesday, 4 April. It is planned to compute 3B41RT and 3B42RT for the affected period, but users should be aware that little microwave data will contribute to 3B42RT. An update will be issued when the recomputed data are posted.

12 February 2007 Upgrades to the power distribution system will require that both TMPA-RT servers be taken down. This is currently scheduled for Tuesday, 20 February in the time window 09:00-15:00 EST (14:00-20:00 UTC). The outage will affect both trmmopen.gsfc.nasa.gov and ftp-tsd.gsfc.nasa.gov.

9 January 2007 A Center-wide upgrade to the network at GSFC caused intermittent loss of input data for all 3B4xRT products during 6-7 January 2007. Most of the data subsequently arrived, so a general reprocessing was carried out for affected times, although some SSM/I data were not retrieved. The reprocessed files placed in the archive are:

3B40RT: 7 January 2007 00Z-15Z
3B41RT: 6 January 2007 14Z-23Z, 7 January 2007 00Z-19Z
3B42RT: 6 January 2007 15Z-21Z, 7 January 2007 00Z-18Z

21 September 2006 The trmmopen.gsfc.nasa.gov anonymous FTP site was taken down for emergency maintenance; a mirror FTP site has been established at ftp-tsd.gsfc.nasa.gov/merged/, which is intended as a long-term alternative for users to access when trmmopen is not responding. Presently the directory structure on ftp-tsd is different - no leading "pub" and no subdirectories for the separate products. It is anticipated that the directory structure on ftp-tsd will be upgraded once new hardware is acquired in early 2007. The current estimate is that trmmopen will be restored by 25-26 September.

12 September 2006 A temporary fix to the AMSR-E feed has been made; AMSR-E is absent from the products for the data times 12Z 07 September to 03Z 11 September 2006.

Belatedly we post here that the anonymous FTP site for the TMPA-RT was moved to [ftp://trmmopen.gsfc.nasa.gov/pub/merged/](http://trmmopen.gsfc.nasa.gov/pub/merged/) in January 2006.

8 September 2006 Two recent events have reduced the microwave data content in 3B40RT and 3B42RT. First, DoD activated a radar calibration beacon on the F15 SSM/I on 14 August that interferes with the 22V channel and effectively prevents reliable estimates with the current version of GPROF. Second, AMSR-E dropped out starting 5 September, first due to access issues, and then to a hardware failure at the AMSR-E data provider machine. It is possible that GPROF will be ready to utilize data from the F16 SSM/IS follow-on to F15 by early October, while the hardware supporting AMSR-E should be fixed by about 13 September. In the interim, users should be aware of the possibility of reduced quality in the estimates.

The primary refereed citation for the TMPA is in press:

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, D.B. Wolff, 2006: The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeor.*, to appear.

5 June 2006 Erroneous GMS data for the span 00Z-13Z on 31 May 2006 forced us to recompute 3B41RT and 3B42RT data for that period. The original estimates were removed and the new posted to the trmmopen ftp site around 14Z on 5 June 2006. This only affects longitudes roughly from southeast Asia to New Zealand.

21 July 2005 A corrupted intermediate file halted processing with the 09Z 19 July 2005 data; the first restart on 20 July resulted in errors in 3B40RT (and so for 3B42RT) for 09-21Z 19 July 2005. This second problem was diagnosed and all affected fields recomputed and reposted within 6 hours.

16 June 2005 In late May 2005 the developers realized that 3B41RT (and so 3B42RT to a lesser extent) episodically displayed unrealistic histograms of rain rate. By 1 June 2005 the problem had been traced to overflows in intermediate accumulation arrays and fixed. Both 3B41RT and 3B42RT were subsequently recomputed in mid-June 2005 for the entire period with questionable data, namely 3 February - 31 May 2005.

10 March 2005 A short-coming in GPROF's treatment of SSM/I input data was corrected. Previous to 10 March 2005 the GPROF-SSM/I estimates used in the MPA-RT would very rarely display a rain area that had the correct shape, but unrealistically high values covering the entire feature. The problem was tracked to a lack of screening for unrealistic values in an interpolation routine, and subsequent processing that caused these values to damage all non-zero rain estimates in a 200-scan block, which is a chunk of the orbit about 22° of lat. in length. Known examples of this problem were all observed over ocean, but it is possible that less-obvious cases occurred over land as well. Very sporadic examples of such errors are embedded in the entire MPA-RT [*now known as "TMPA-RT"*] record before 10 March 2005; 3B40RT and 3B42RT were directly affected because they incorporate the GPROF-SSM/I estimates, while 3B41RT was slightly affected to the extent that the occasional erroneous precipitation area was pooled with a month of other, normal estimates for the calibration. This short-coming is believed to affect all versions of GPROF, but only becomes a problem when the input data contain bad values. In particular, the Version 6 GPROF-SSM/I estimates are computed from the Remote Sensing Systems version of Tb's, which contains significantly stronger quality controls than the FNMOC SDR Tb's used in the MPA-RT.

11 February 2005 Upgrades to the computing system at CPC temporarily halted the transmission of CPC Merged 4-Km IR data starting with 14Z 8 February 2005. Thereafter, administrative permission issues delayed resumption of data delivery to TSDIS until about 15Z 10 February 2005. Input data were stored at CPC and complete IR fields were generated when deliveries resumed. This hold-up resulted in suspension of the 3B41RT and 3B42RT processing until the input became available. As a separate issue, the trigger time for computing the MPA-RT [*now the "TMPA-RT"*] has been increased as we study the latency of the new input products. We expect to query the user community about the "best" trigger time to balance data availability and MPA-RT timeliness.

3 February 2005 Version 1.3 was instituted as of 08Z 3 February 2005, providing several important upgrades:

- Precipitation estimates from the AMSR-E and the 3 AMSU-B sensors are incorporated in the HQ product (3B40RT), which nearly doubles the coverage by HQ data in the latitude band 50°N-S from ~45% to nearly 80%.
- Inter-satellite calibration in the HQ product is climatological, reducing the real-time computational load and preparing for the eventual termination of TRMM. The calibrations of AMSU-B and AMSR-E to TMI each have one set of coefficients for land and a separate set for ocean, while SSM/I uses one set for land and 5 for ocean, covering the latitude bands 90-30°S, 30-10°S, 10°S-N, 10-30°N, and 30-90°N. AMSR-E uses a 2-month set of match-ups to ensure sufficient sampling, while all of the others work with single-month accumulations. The AMSR-E and AMSU-B coefficients apply to the entire year, while SSM/I uses a separate set for each season.
- VAR coefficients are recomputed every 3 hr to better control unrealistically high VAR estimates when a batch of unusually cold IR Tb's are encountered in a region. The calculation is now done with all of the HQ-IR match-ups observed in the previous 5 pentads, plus whatever match-ups have occurred in the current pentad.

21 January 2005 Data problems at CPC created an outage in 3B41RT and 3B42RT for the period 00Z 26 December to 15Z 27 December 2004. After the problems were corrected and the CPC 4-km reposted, 3B41RT and 3B42RT were computed and reposted on 30 December 2004.

4 January 2005 The real-time GPROF-TMI (2A12RT) was upgraded to Version 6 at 1754Z.

23 December 2004 A disk space issue and data errors in the original CPC 4-km IR data combined to create an outage in 3B41RT and 3B42RT for the period 12Z 17 December to 12Z 20 December 2004. After the problems were corrected and the CPC 4-km reposted, 3B41RT and 3B42RT were computed and reposted on 22 December 2004.

16 November 2004 Errors in creating GOES sectors within NOAA were included in the original CPC 4-km IR data and hence in 3B41RT and 3B42RT for the period 17Z 9 December to 06Z 11 December 2004. After the sectors were corrected and the CPC 4-km reposted, 3B41RT and 3B42RT were recomputed (with calibration coefficients updated on 00Z 11 December 2004) and reposted on 16 December 2004.

23 November 2004 A gap in real-time processing occurred on 21-22 August 2004 that was considered unacceptably long. The 3B40RT files were not easily recoverable, but 3B41RT files for the date/times

21 August 2004: 06-13, 15, 18, 21

22 August 2004: 00, 07-10, 12, 15

were computed using the calibration file created 31 August 2004, and 3B42RT files for the date/times

21 August 2004: 06, 09, 12, 15, 18, 21

22 August 2004: 00, 06, 09, 12, 15

were created using only 3B41RT as input. These results were posted on 19 November 2004.

15 April 2004 The calibration of microwave estimates to the TMI was upgraded, starting with the 00Z 6 April 2004 data:

- Zeroes are now included in the calibration histograms, improving the treatment of fractional coverage.
- The calibrations are adjusted so that each microwave estimate reproduces the precipitation volume of the matched TMI. This should boost the total rain volume slightly, as some sensors have a fractional coverage by rain that is less than the TMI's.
- A second ambiguous pixel screening was introduced to control a residue of artifacts, mostly in high polar latitudes.

The calibration of IR Tb's to the HQ estimates was upgraded, starting with the 04Z 15 April 2004 data:

- IR coefficients are no longer smooth-filled across areas lacking HQ data (mostly due to cold/icy/frozen surface-induced drop-outs). Instead, IR-coincident accumulated precipitation and fractional coverage by precipitation are smooth-filled, then used to derive the IR coefficients. This should improve instantaneous fields in cold land areas substantially, although the resulting accumulations will continue to have high uncertainty.
- The coldest-Tb rain rates are no longer set to a constant (determined locally by match-ups with the HQ). Instead, a climatological Tb-rainrate curve is applied with the rainrates adjusted to be piecewise continuous to the curve determined from the matchups. This should alleviate cold-Tb areas showing constant rainrates.

12 February 2004 GPROF-SSM/I Version 6.5 was installed, which features the ocean code from Version 6 and the land and coast code from Version 7. We chose not to rework the TMI-SSM/I calibration coefficients, so estimates over land and coast (and the corresponding IR calibration) will be somewhat less accurate for the following 2-week period, most likely having a small negative bias that improves with each additional pentad of Version 6.5 data.

10 October 2003 An outage at FNMOC resulted in the following loss of SSM/I data:

F13: 20031008/09:55 - 09/19:39 UTC

F14: 20031008/10:16 - 09/18:39 UTC

F15: 20031008/11:20 - 09/19:24 UTC

As a result 3B40RT for 20031008/12 - 10/15 UTC only contains TMI data, resulting in highly reduced data coverage. There is no affect on 3B41RT, and relatively little on 3B42RT.

25 September 2003 Hurricane Isabel caused significant interruptions in receiving and processing the requisite IR input data, which precluded the computation of 3B41RT and 3B42RT estimates during parts of the days 16-22 September 2003. After real time the necessary IR input data became available and 3B41RT and 3B42RT were computed for the entire period of 18Z/16 - 12Z/22 September 2003. The microwave-based IR calibration coefficients used for the post-real-time computation were those generated on 00Z/23 September 2003, while those few files generated in real time were based on either the 00Z/13 or 00Z/18 September 2003 coefficient sets (for times before or from 00Z/18, respectively). Comparisons of a few recomputed files with available files computed in real time showed generally small differences, but a few were locally non-trivial. Consequently, for consistency the entire sequence of sparse real-time files for the period 18Z/16 - 12Z/22 September 2003 were replaced by the post-real-time files computed using the "new" coefficients.

20 June 2003 The sample file-reading routine in IDL was corrected to size the header at 2880 bytes. The previous incorrect value of 1440 bytes made the data field appear to be translated by 180° of long.

2 March 2003 Starting 09Z 28 February 2003 the real-time production system was upgraded to more effectively eliminate artifacts in the microwave estimates over frozen and snowy surfaces. Starting 00Z 02 March 2003 the upgrade was extended to the microwave calibration of the infrared (IR) estimates. These changes improve all three products, since the microwave and IR estimates are used to create the final combination. The artifacts mainly appeared in Eurasia north of 35°N.

Introduction

The system to produce the "TRMM and Other Data" estimates in real time was developed to apply new concepts in merging quasi-global precipitation estimates and to take advantage of the increasing availability of input data sets in near real time. The overall system is referred to as the real-time TRMM Multi-Satellite Precipitation Analysis (TMPA-RT). The TMPA-RT is run quasi-operationally on a best-effort basis at the TRMM Science Data and Information System (TSDIS), with on-going scientific development by the research team led by Dr. Robert Adler in the GSFC Laboratory for Atmospheres. Estimates are posted to the web about 6 hours after observation time, although processing issues may delay or prevent this schedule. Due to the experimental nature of these estimates, users are encouraged to report their experiences with the data, and they should expect episodic upgrades or outages as the system develops.

Product Definitions

There are three "TRMM and Other Data" products:

3B40RT (High Quality, or HQ)

The 3B40RT combined microwave precipitation estimate provides a global 0.25°x0.25°-averaged 3-hourly combination of all available TMI, SSM/I, AMSR-E, and AMSU-B estimates:

1. Offline, the GPROF-SSM/I, GPROF-AMSR, and AMSU-B have been probability-matched to 2A12RT. Prior to 08Z 3 February 2005, this was done routinely for GPROF-SSM/I as 2 sets, one for ocean and one for land, computed every 5 days from the last 30 days of SSM/I-IR match-ups. Thereafter, the calibrations of AMSU-B and AMSR-E to TMI each have one set of coefficients for land and a separate set for ocean, while SSM/I uses one set for land and 5 for ocean, covering the latitude bands 90-30°S, 30-10°S, 10°S-N, 10-30°N, and 30-90°N. AMSR-E uses a 2-month set of match-ups to ensure sufficient sampling, while all of the others work with single-month accumulations. The AMSR-E and AMSU-B coefficients apply to the entire year, while SSM/I uses a separate set for each season.

2. The GPROF-SSM/I, GPROF-AMSR, AMSU-B, and 2A12RT estimates are gridded to a $0.25^\circ \times 0.25^\circ$ grid for a 3-hour period centered on the major synoptic times (00Z, 03Z, ..., 21Z). Prior to 08Z 3 February 2005, only GPROF-SSM/I and 2A12RT estimates were used.
3. The GPROF-SSM/I, GPROF-AMSR, and AMSU-B estimates are calibrated to 2A12RT.
4. The rain rate in each grid box is the pixel-weighted average of 2A12RT, GPROF-SSM/I, and GPROF-AMSR grid boxes contributing during the 3 hours, or the pixel-weighted average of AMSU-B if no other HQ estimates are available.
5. Additional fields in the data file include the number of pixels, the number of pixels with non-zero rain, and the number of pixels for which the estimate is "ambiguous," or highly uncertain.
6. The SSM/I and AMSR-E data are available in the latitude band 85°N-S , but GPROF only returns estimates in the band 70°N-S . AMSU-B data are available over the entire globe, and estimates are attempted in the entire domain. However, all of the HQ algorithms are unable to provide estimates in regions with frozen or icy surfaces.
7. In a future upgrade the random error will be estimated. Currently the random error field is set to missing.

3B41RT (Variable Rainrate, or VAR)

The 3B41RT IR precipitation estimate converts $0.25^\circ \times 0.25^\circ$ -averaged geo-IR Tb to rainrates that are HQ-calibrated locally in time and space:

1. Both geo-IR Tb and HQ are averaged to $0.25^\circ \times 0.25^\circ$ to ensure consistent spatial scale, and time-space matched data are accumulated over calendar months.
2. In each calibration, the Tb-rainrate curve is set locally by probability matching the month's histograms of coincident IR Tb and HQ rain rate.

The local VAR Tb-rainrate curve is applied to each Merged 4-Km IR Tb data set in the month:

1. Over most of the globe the on-hour data field is taken as the input data, with fill-in by the previous half-hour image. The exception is the GMS sector, where the previous half-hour is primary, since GMS does not schedule images on the hour. [In that case, much of the GMS sector is filled with data from METEOSAT5 and GOES-W at very high zenith angles.]
2. The Tb-to-rainrate conversion is a simple look-up, using whatever set of VAR calibration coefficients is current.
3. In a future upgrade the random error will be estimated. Currently the random error field is set to missing.

3B42RT (Combination of HQ and VAR)

The 3B42RT combination is computed every 3 hours from that hour's HQ and VAR fields:

1. The present combination scheme is to take the HQ field wherever it is non-missing, and fill in with VAR elsewhere.
2. The additional field in the file is the RMS error of the estimate.
3. The VAR estimates are only posted for the latitude band 50°N-S .

4. It is planned to do a more sophisticated combination in a future release.
5. Following the computation of the monthly SG combination (see that topic for details), in each grid box all of the available 3-hourly HQ+VAR values are scaled to sum to the monthly SG value.
6. In a future upgrade the random error will be estimated. Currently the random error field is set to missing.

File Contents and Format

Table 1. File layout for 3B40RT, 3B41RT, 3B42RT.

	<i>3B40RT</i>		<i>3B41RT</i>		<i>3B42RT</i>	
<i>Block</i>	<i>Byte Count</i>	<i>Field</i>	<i>Byte Count</i>	<i>Field</i>	<i>Byte Count</i>	<i>Field</i>
1	2880	header	2880	header	2880	header
2	2073600*	precip	1382400&	precip	1382400&	precip
3	2073600*	error	1382400&	error	1382400&	error
4	1036800+	# pixels	691200@	# pixels	691200@	source
5	1036800+	# ambig. pixels	-	-	-	-
6	1036800+	# rain pixels	-	-	-	-

* INTEGER*2, 90°N-S & INTEGER*2, 60°N-S

+ INTEGER*1, 90°N-S @ INTEGER*1, 60°N-S

Header

Each file starts with a header that is one 2-byte-integer row in length, or 2880 bytes. The header is ASCII in a "PARAMETER=VALUE" format that makes the file self-documenting (e.g., "algorithm_id=3B40RT"). As such, the header can be read with standard text editors, output as text with simple application programs, or parsed for input into applications. Successive "PARAMETER=VALUE" sets are separated by spaces, and no spaces or "=" are permitted in either PARAMETER or VALUE. The current PARAMETER entries and definitions are:

PARAMETER	Definition
algorithm_ID	TRMM algorithm identifier (e.g., "3B40RT")
algorithm_version	Version of the science algorithm
granule_ID	TSDIS granule identifier (e.g., "3B40RT.2001121809.bin")
header_byte_length	Number of bytes in the header
file_byte_length	Number of bytes in the file, expressed as a formula describing the file structure
nominal_YYYYMMDD	Nominal UTC year, month, and day of the month
nominal_HHMMSS	Nominal UTC hour, minute, and second
begin_YYYYMMDD	Start UTC year, month, and day of the month
begin_HHMMSS	Start UTC hour, minute, and second
end_YYYYMMDD	End UTC year, month, and day of the month
end_HHMMSS	End UTC hour, minute, and second
creation_YYYYMMDD	Date the file was created as year, month, and day of the month
west_boundary	Longitude of the western edge of the data domain

east_boundary	Longitude of the eastern edge of the data domain
north_boundary	Latitude of the northern edge of the data domain
south_boundary	Latitude of the southern edge of the data domain
origin	Geographical direction of the first grid box from the grid center
number_of_latitude_bins	Number of grid boxes in the meridional direction
number_of_longitude_bins	Number of grid boxes in the zonal direction
grid	Size of one grid box
first_box_center	Geolocation of the first grid box center
second_box_center	Geolocation of the second grid box center
last_box_center	Geolocation of the last grid box center
number_of_variables	Number of data fields
variable_name	List of the data field names, separated by commas
variable_units	List of data field units, separated by commas, in the same order as the variable_name list
variable_scale	List of data field scaling factors, separated by commas, in the same order as the variable_name list
variable_type	List of data field word types, separated by commas, in the same order as the variable_name list
byte_order	Order of bytes in a data word ("big_endian" or "little_endian")
flag_value	List of special values, separated by commas
flag_name	List of special value names, separated by commas, in the same order as the flag_value list
contact_name	Name of the person to contact with questions
contact_address	Postal address of the contact_name
contact_telephone	Telephone number of the contact_name
contact_facsimile	Facsimile number of the contact_name
contact_email	Email address of the contact_name

Thereafter the data fields follow. All the fields are on a 0.25° lat./lon. grid that increments most rapidly to the east (from the Prime Meridian) and then to the south (from the northern edge). Grid box edges are on multiples of 0.25°. The data fields are written as flat binary data in big-endian byte order.

3B40RT

Following the header, 5 data fields appear:

precipitation	(2-byte integer)
precipitation_error	(2-byte integer)
total_pixels	(1-byte integer)
ambiguous_pixels	(1-byte integer; highly uncertain values)
rain_pixels	(1-byte integer)

All fields are 1440x720 grid boxes (0-360°E,90°N-S). The first grid box center is at (0.125°E,89.875°N). Files are produced every 3 hours on synoptic observation hours (00 UTC, 03 UTC, ..., 21 UTC) as an accumulation of all HQ swath data observed within +/-90 minutes of the nominal file time. Estimates are only computed for the band 70°N-S.

3B41RT

Following the header, 3 data fields appear:

Precipitation (2-byte integer)
precipitation_error (2-byte integer)
total_pixels (1-byte integer)

All fields are 1440x480 grid boxes (0-360°E,60°N-S). The first grid box center is at (0.125°E,59.875°N). Files are produced every hour from the on-hour IR image (except for the previous half-hour image for GMS), with fill-in by the previous half-hour image (except for GMS, where the on-hour image is used for fill-in). Valid estimates are only provided in the band 50°N-S.

3B42RT

Following the header, 3 data fields appear:

Precipitation (2-byte integer)
precipitation_error (2-byte integer)
source (1-byte integer; 0,1,2,4 stand for none, HQ, VAR, and sparse HQ)

All fields are 1440x480 grid boxes (0-360°E,60°N-S). The first grid box center is at (0.125°E,59.875°N). Files are produced every 3 hours on synoptic observation hours (00 UTC, 03 UTC, ..., 21 UTC) using that hour's 3B40RT and 3B41RT data sets. Valid estimates are only provided in the band 50°N-S.

Note that we use the term "gridbox" to denote the values on Level 3 data (i.e., gridded data), while we use the term "pixel" to denote individual values of Level 2 data (i.e., instrument footprints). Thus, there can be many pixels contributing to a gridbox.

Both precipitation and random error are scaled by 100 before conversion to 2-byte integer. Thus, units are 0.01 mm/h. To recover the original floating-point values in mm/h, divide by 100. Missings are given the 2-byte-integer missing value, -31999. The remaining fields are in numbers of pixels, except the source variable, which is dimensionless.

Currently the random error fields are all set to the 2-byte-integer missing value, -31999. This placeholder will be replaced with actual estimates as development proceeds.

The variable ambiguous_pixels is the count of pixels for which the algorithm cannot determine whether the scene has valid or invalid data. It is a subset of the total_pixel and many, but not all, are included in raining_pixels. In general, a "high" fraction of ambiguous_pixels indicates that the grid box value is invalid.

The originating machine on which the data files are written is a Silicon Graphics, Inc. Unix workstation, which uses the "big-endian" IEEE 754-1985 representation of 4-byte floating-point unformatted binary numbers. Some CPUs, including PCs and DEC machines, might require a

change of representation (i.e., byte swapping) before using the data. In some cases, the gunzip routine, used to uncompress the data, will change representations automatically.

Special Values

All of the scaled 2-byte-integer precipitation and random error fields have one value with special meaning. Any grid box with insufficient valid data to make an estimate is assigned the 2-byte-integer value -31999. As well, the scaled 2-byte-integer precipitation and random error fields are clipped to [-31998,31998] to prevent duplication of the missing value (at the negative end) or overflows (at both the positive and negative ends). Note that any examples of clipping should be immediately reported to the dataset developers.

In addition, 3B40RT precipitation values that are highly likely to be artifacts (ambiguous fraction of pixels at least 40% or 5x5-gridbox averaged ambiguous fraction at least 20%) are encoded as $(-p - 0.01)$, where "p" is the original precipitation value, before conversion to scaled 2-byte-integer. Thus, users can recover the estimated value of such gridboxes if desired, but the usual scheme of requiring precipitation to be non-negative will filter out these suspect values.

Likewise, 3B41RT and 3B42RT precipitation values outside the 50°N-S latitude band are considered experimental and are encoded as $(-p - 0.01)$, where "p" is the original precipitation value, before conversion to scaled 2-byte-integer. Thus, users can recover the estimated value of such gridboxes if desired, but the usual scheme of requiring precipitation to be non-negative will filter out these suspect values.

The 3B42RT "source of estimate" field only has three discrete values, -1, 0, 100, which correspond to "no estimate", "HQ", and "VAR".

Note that any negative values in the various "number of" fields is a processing error that should be immediately reported to the dataset developers.

Dataset Validation

Formal validation studies are underway. The time series of the global images shows good continuity in time and space across the geo-IR data boundaries. Overall, the analysis approach appears to be working as expected. See Huffman et al. (2007) for more information.

Early validation studies are being conducted under the auspices of the International Precipitation Working Group in Australia, the continental U.S. and western Europe. Respectively, the web sites for these activities are:

http://www.bom.gov.au/bmrc/SatRainVal/sat_val_austr.html
http://www.cpc.ncep.noaa.gov/products/janowiak/us_web.shtml
<http://kermit.bham.ac.uk/~ipwgeu/>

The primary limitations on the HQ (3B40RT) are the sparse sampling by the collection of passive-microwave satellites and algorithm drop-outs in regions with icy or frozen surface. The infrared results (3B41RT) are designed to emulate the microwave results as closely as possible, so known deficiencies in the microwave will likely be reflected in the infrared as well. In addition, it is well-known that infrared algorithms of the kind used here have large random errors at the fine time and space scales provided. However, we expect the infrared estimates to match the histogram of microwave estimates, so that user-specified averaging should yield approximately unbiased results. Finally, the combined microwave-IR fields (3B42RT) contain data boundaries between the regions of microwave and IR coverage. Instantaneously the boundaries are usually subtle, but are more noticeable in movie loops, since the regions of coverage change with each image. We encourage users to report successes and problems in applying these datasets to their particular applications.

Dataset Status

Beta testing began in early December 2001. An official (experimental) version was instituted in late January 2002. Processing changes occurred on 6 February and 12 March 2002, and subsequent developments are documented in the "Latest News" and "Previous News" at the top of this document. Users should anticipate a series of versions as the algorithm is developed further. The present areas of interest are: calibrating the RT to be approximately unbiased with respect to the Version 6; improving the HQ product by auditing out AMSU-B data that are deficient in precipitation coverage; and moving to shorter-interval estimation periods to more accurately represent the time series of precipitation.

Currently, access to the digital data is by anonymous FTP to [trmmopen.gsfc.nasa.gov](ftp:trmmopen.gsfc.nasa.gov). Under subdirectory `pub/merged` there are three directories:

- CombinedMicro (3B40RT)
- CalibratedIR (3B41RT)
- MergeIRMicro (3B42RT)

There is also a mirror site at <ftp:tsdis.gsfc.nasa.gov>, with a single flat directory "merged".

Example Programs

The data fields are all written with C-language code as blocks of bytes, so there are no extraneous bytes in the files. Because the first two fields are 2-byte integers and the rest are 1-byte integers in each file (to save space), users must exercise care in using FORTRAN direct access to read the data. The FORTRAN example programs read all fields with a single OPEN. Alternatively, the files can be opened with different logical record sizes depending on whether one is reading 2-byte-integer or 1-byte-integer fields. Note well that the units of the logical record size is not part of the FORTRAN 77 standard. On SGI machines it is in 4-byte words, but some other systems expect it in bytes. Also, to repeat an earlier comment, the originating machine on which the data files are written is a Silicon Graphics, Inc. Unix workstation. It uses the "big-endian" IEEE 754-1985 representation of 4-byte floating-point unformatted binary

numbers, and some CPUs, such as PCs and DEC machines, might require a change of representation (i.e., byte swapping) before using the data.

The FTP site <ftp://trmmopen.gsfc.nasa.gov/pub/merged/software> provides several example programs:

<code>idlsave.pro</code>	IDL journal file showing output of a particular 3B40RT header
<code>read3B4XRT.c</code>	C example
<code>read_header.f</code>	FORTTRAN header-read example
<code>read_rt_file.f</code>	FORTTRAN single-read example
<code>read_rt_file.pro</code>	IDL example
<code>read_rt_lines.f</code>	FORTTRAN line-by-line example

These are also available at ftp://meso.gsfc.nasa.gov/pub/agnes/huffman/rt_examples/rt_docs.

Example Images and Movies

Users may obtain example GIF images and QuickTime movies at <http://trmm.gsfc.nasa.gov>. In addition, users may create their own subsets, averages, and time series, downloadable as plots or ASCII data from the TRMM Online Visualization System (TOVAS), at <http://disc2.nascom.nasa.gov/Giovanni/tovas/>. TOVA is provided by the Goddard Earth Science Data and Information Services Center (GES DISC).

Additional Documentation

Users should refer to the detailed documentation (3B4XRT.pdf) and programming examples at the anonymous FTP site <ftp://trmmopen.gsfc.nasa.gov/pub/merged/software> for additional details.

Primary Reference

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, D.B. Wolff, 2007: The TRMM Multi-satellite Precipitation Analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeor.*, **8**(1), 38-55. PDF available at ftp://meso.gsfc.nasa.gov/agnes/huffman/papers/TMPA_jhm_07.pdf.gz